Identification of growth-coupled production strains considering protein costs and kinetic variability

Dinh, Hoang V.; King, Zachary A.; Palsson, Bernhard O.; Feist, Adam M.

Published in:
Metabolic Engineering Communications

Link to article, DOI:
10.1016/j.mec.2018.e00080

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Metabolic Engineering Communications 7 (2018) e00080

Contents lists available at ScienceDirect

Metabolic Engineering Communications

journal homepage: www.elsevier.com/locate/mec

Identification of growth-coupled production strains considering protein costs and kinetic variability

Hoang V. Dinh[a], Zachary A. King[a], Bernhard O. Palsson[a,b,c], Adam M. Feist[a,c,*]

[a] Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive #0412, La Jolla, CA 92093-0412, USA
[b] Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive #0412, La Jolla, CA 92093-0412, USA
[c] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, DK-2800 Kongens, Lyngby, Denmark

ABSTRACT

Conversion of renewable biomass to useful molecules in microbial cell factories can be approached in a rational and systematic manner using constraint-based reconstruction and analysis. Filtering for high confidence *in silico* designs is critical because *in vivo* construction and testing of strains is expensive and time consuming. As such, a workflow was devised to analyze the robustness of growth-coupled production when considering the biosynthetic costs of the proteome and variability in enzyme kinetic parameters using a genome-scale model of metabolism and gene expression (ME-model). A collection of 2632 unfiltered knockout designs in *Escherichia coli* was evaluated by the workflow. A ME-model was used in the workflow to test the designs' growth-coupled production in addition to a less complex genome-scale metabolic model (M-model). The workflow identified 634 M-model growth-coupled designs which met the filtering criteria and 42 robust designs, which met growth-coupled production criteria using both M and ME-models. Knockouts were found to follow a pattern of controlling intermediate metabolite consumption such as pyruvate consumption and high flux subsystems such as glycolysis. Kinetic parameter sampling using the ME-model revealed how enzyme efficiency and pathway tradeoffs can affect growth-coupled production phenotypes.

1. Introduction

The chemical industry has relied on petroleum as raw material for the last century (Sittig and Weil, 1954). Shifting to renewable biomass as feedstocks has gained interest in both industry and academia as a long term solution for feedstocks (Johnson, 2008; Lee and Kim, 2015). Microbial cell factories can be used in efficient bioprocesses to convert biomass to a wide array of useful products (Lee and Kim, 2015). Microbial cell factories need to be designed to optimize for production rate, yield, and titer as wild-type strains do not generally produce desirable molecules. For example, in the model bacterium *Escherichia coli*, CO_2 is produced aerobically and a mixture of formate, ethanol, acetate, D-lactate, and succinate are excreted anaerobically (Clark, 1989). Knocking out (*i.e.,* removing through genetic manipulation) the default fermentation pathways is a common strategy that redirects raw material toward desirable products (King et al., 2017). Byproduct excretion during optimal cell growth, called growth-coupled production, is a desirable target for strain design as adaptive laboratory evolution can be used to enhance growth-coupled production by selecting for cells with higher fitness (Fong et al., 2005; Zhang et al., 2007).

Fueled by advances in systems biology, synthetic biology, and evolutionary engineering, rational metabolic engineering has been successfully applied to the design of microbial cell factory strains (Davy et al., 2017; Lee and Kim, 2015; Park and Lee, 2008). A systems biology approach with strain design evaluation in genome-scale models (GEMs) takes into account multiple biological components and their interactions that are necessary for predicting growth-coupled production of target molecules. GEMs are collections of genetic and biochemical information from databases and literature (Thiele and Palsson, 2010). Constraint-based reconstruction and analysis (COBRA) methods can be used with GEMs to calculate metabolic flux distributions and predict genotype-phenotype relationships (Lewis et al., 2012). Systematic approaches *via* optimization algorithms (Machado and Herrgård, 2015; Maia et al., 2016) can be performed with GEMs to identify strain designs to optimize cell factory strains. In the recent years, models of metabolism and gene expression (ME-models) have been reconstructed with additional biological constraints, allowing for more accurate predictions of overflow metabolism (O'Brien et al., 2013), membrane content (Liu et al., 2014), and by-product secretion (King et al., 2017).

* Corresponding author at: Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive #0412, La Jolla, CA 92093-0412, USA.
E-mail address: afeist@ucsd.edu (A.M. Feist).

https://doi.org/10.1016/j.mec.2018.e00080
Received 26 January 2018; Received in revised form 25 September 2018; Accepted 7 October 2018

Model-driven strain design and pathway prediction methods (Campodonico et al., 2014; Feist et al., 2010) provide an ample number of *in silico* strain designs, but it is currently infeasible to test all of these designs *in vivo*. Therefore, methods to filter and refine *in silico* strain design are needed. Growth-coupled strain designs are susceptible to alternative production phenotypes in which undesired byproducts are excreted in place of the target molecule without a significant decrease in growth rate. Using an M-model (*i.e.,* a metabolic model), susceptibility to alternative production phenotypes can be evaluated by minimizing production of the target molecule *in silico* at the maximum growth rate. Additionally, changing ME-model kinetic parameters has been shown to affect alternative production phenotypes by changing the protein costs of competing fermentation pathways (King et al., 2017). Thus, to add confidence for growth-coupled designs, strain designs can be additionally tested with a ME-model kinetic parameter sampling approach. By tweaking the kinetic parameters and simulating cell growth in the ME-model, target growth-coupled production can be tested under different scenarios of enzyme and pathway efficiency.

This work presents an efficient, high-throughput workflow to filter a collected pool of 2632 *in silico E. coli* strain designs from previous studies to identify high-confidence, *robust* designs. *Robust* strain designs are predicted to have growth-coupled production across a range of turnover rate (k_{eff}) parameter values in the ME-model. The first stage of the workflow utilized *E. coli* M-models to identify 634 *significant growth-coupled production* designs that have > 10% carbon yield (Table 1 and Section 2.4) and satisfy additional criteria on the effectiveness of reaction knockouts. The second stage of the workflow tested growth-coupling under kinetic parameter sampling using a ME-model to identify 42 high-confidence designs. Enzyme efficiency was shown to be a decisive factor for the presence of growth-coupled production in the ME-model. *Robust* strain designs with predicted growth-coupled production of target molecules in both M- and ME-models are suggested for experimental implementation and *in vivo* testing.

2. Methods

2.1. Knockout strain design pool

A total of 9 growth conditions were tested by combining 3 primary carbon substrates (glucose, xylose, and glycerol) with 3 types of oxygenation states (aerobic, ECOM, and anaerobic). These substrates and environmental parameters were selected based on previously established criteria (Campodonico et al., 2014; Feist et al., 2010). The *E. coli* cytochrome oxidase mutant (ECOM) strain which has an "aerobic fermentation" phenotype is simulated in an aerobic oxygenation state with knockouts of the cytochrome oxidase reactions CYTBDpp, CYTBD2pp, and CYTBO3_4pp (Portnoy et al., 2008).

To differentiate the design pools generated at different steps in the workflow (Section 3.1), the three major design pools were hereby labeled and mentioned throughout this work. The initial global pool of growth-coupled designs, hereby called '*unfiltered*' designs, was collected from previous studies. After workflow steps 1 and 2 (Fig. 2), the designs that had growth-coupled production with carbon yield > 10% (Table 1 and Section 2.4) in the M-model and had redundant knockouts removed were called '*significant growth-coupled production*' or simply '*significant production*' designs. After workflow steps 3 and 4 (Fig. 2), the designs which maintained their target growth-coupled production under ME-model kinetic parameter sampling were called '*robust*' growth-coupled designs. *Robust* designs were so labeled because their growth-coupled production was present across simulations, including M-model and ME-model with multiple kinetic parameters sets. The list of designs passing each filter steps were provided in the Supplementary data.

2.2. Models

The *unfiltered* design pool was obtained from previous studies that implemented strain design algorithms in the *i*AF1260b (1668 metabolites, 2388 reactions, and 1261 genes) and *i*JO1366 (1805 metabolites, 2583 reactions, and 1367 genes) M-models of *E. coli* K-12 MG1655 (Feist et al., 2010, 2007; Orth et al., 2011) (See Table 1 for definitions of terms used in this study). M-model simulations were performed to evaluate the carbon yield of production (Section 2.4 and Fig. 2's step 1) and remove redundant knockouts (Section 2.5 and Fig. 2's step 2) in the corresponding model in which the strain design was found. The latest *E. coli* ME-model *i*LE1678-ME (Lloyd et al., 2018) was used for design robustness evaluation (Section 2.6 and Fig. 2's steps 3 and 4). The ME-model *i*LE1678-ME (7031 metabolites, 12,654 reactions, and 1678 genes) has the capability to explicitly calculate the cost of gene expression machinery supporting the metabolic flux state and thus accounts for both metabolic flux distribution and gene expression machinery cost in optimization.

Table 1
Table of terms and their definitions used in the work.

Term	Definition
Nonphysiological pathways	Pathways which were identified to not be active *in vivo* and thus cause unobserved phenotypes during *in silico* simulation. Literature evidence was found to support their removal during simulations (Supplementary text, Section S1)
Carbon yield ($y_{product}$)	Yield calculated by fraction of carbon from substrate converted into target molecule (mmol carbon in product / mmol carbon in substrate)
Significant growth-coupled production	A design has *significant growth-coupled production* or so called a *significant production* design when its minimum target molecule production's carbon yield is more than 10% at maximum growth rate
Maximum growth rate without target molecule production ($\mu_{max,RP}$)	Shutting down an exchange reaction for a target molecule secretion causes the *in silico* cell to find an alternative phenotype with lower maximum growth rate $\mu_{max,RP}$ if possible. This growth-coupled trait was derived from "strength-of-coupling" in Feist et al. (2010).
Substrate-specific productivity (*SSP*)	Multiplying maximum growth rate with $y_{product}$ gives SSP. This growth-coupled trait was derived from "substrate-specific productivity" in Feist et al. (2010).
Redundant knockouts	Removal of redundant knockouts from the design does not decrease carbon yield by 1%, SSP by 1% of the original value, and μ_{RP} by 1% of the original value.
Duplicates	Designs with the exact same set of knockouts, substrate, oxygenation state, target molecule, and heterologous pathway were duplicates. Only one among the duplicates went through the filter.
Turnover rate (k_{eff}) (in ME-model)	Turnover rate (k_{eff}) is a parameter associated with each enzyme in the ME-model. A set of constant k_{eff}s are determined prior to each ME-model simulation. Different sets of k_{eff}s can result in different phenotypes at optimal growth rate.
Robust and non-robust (in ME-model)	*Robust* designs, first of all, were growth-coupled in M-model. Then, designs whose flux through biomass objective function (BOF) was positive accompanied by target molecule production (growth-coupled production) or flux through BOF was zero for all sampled sets of k_{eff}s were *robust*. Designs whose flux through BOF was positive without target molecule produced for at least one sampled set of k_{eff}s were non-robust.

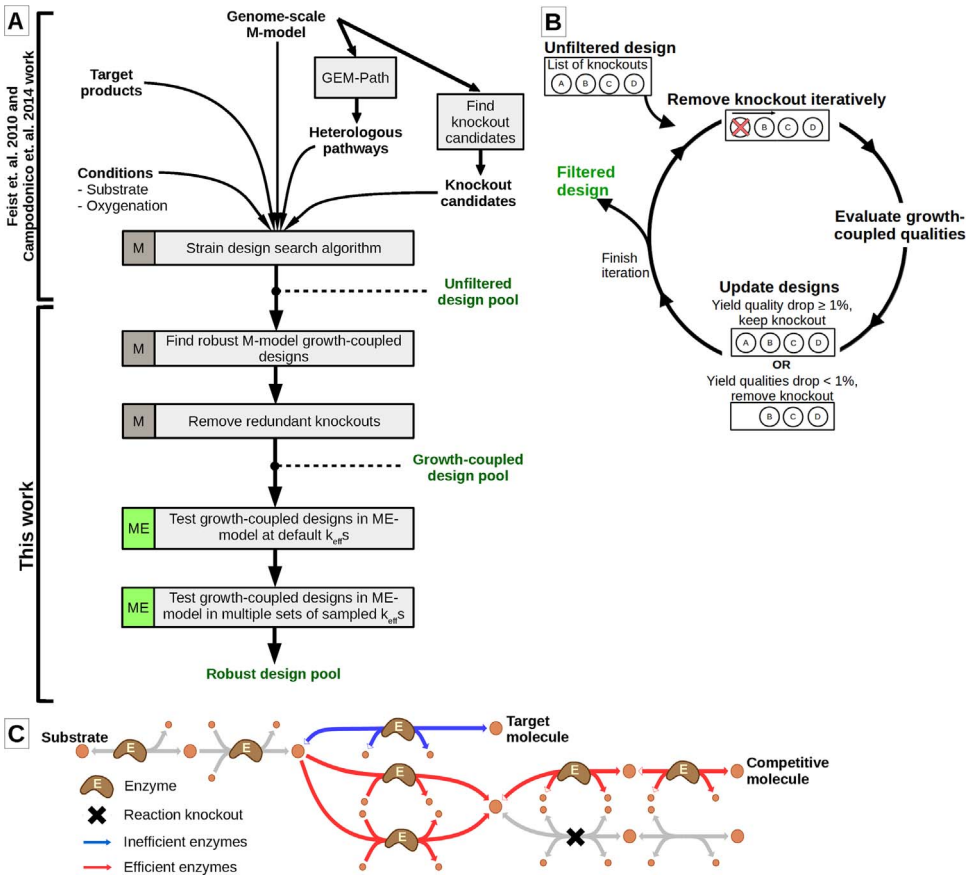

Fig. 1. Growth-coupled production strain design evaluation workflow. (A) Evaluation workflow using M- and ME-models. (B) Iterative workflow to remove redundant knockouts when the removal did not drop the growth-coupled production qualities more than 1%. (C) A schematic example of the effect of enzyme and pathway selection for kinetic parameter sampling in a ME-model. The parameters examined, in some cases, made the target molecule pathway more inefficient and/or a competitive molecule pathway more efficient, thus shifting the optimal phenotypic profile at maximum growth rate.

2.3. Cell growth simulation

Flux balance analysis (FBA) was used to simulate feasible flux distributions. Cell growth was simulated by maximizing the biomass objective function in the M-model and biomass dilution in the ME-model. For M-models, FBA was employed in COBRApy (Ebrahim et al., 2013) and Python 2.7. In M-models, linear optimization problems were solved by the Gurobi solver (Gurobi Optimization Inc., Houston, TX, USA). For ME-model calculations, linear optimizations were solved by the solveME (Yang et al., 2016) software package and the Quad MINOS solver (Ma and Saunders, 2014). The COBRAme toolbox (Lloyd et al., 2018) was used to execute FBA in the ME-model by running a binary search for optimal growth across a range of linear optimizations (as described previously in O'Brien et al., 2013).

Substrate uptake rate was set to 20 mmol gDW^{-1} h^{-1} for M-model simulations. Oxygen uptake rate was set to 20 mmol gDW^{-1} h^{-1} for aerobic conditions and 0 for anaerobic conditions for M-model simulations. Unlimited uptake of substrate and oxygen in aerobic condition were used in ME-model simulations. In an M-model, the substrate uptake is usually constrained with experimental data because the growth rate can increase proportionally and infinitely to the uptake rate. In a ME-model, the substrate uptake rate constraint is unnecessary since proteome limitation will constrain available protein levels in general. Substrate uptake rate is only constrained to a low value if the goal is to simulate a nutrient-limited scenario (O'Brien et al., 2013). Reaction knockouts were simulated by constraining the flux of the corresponding reactions to be zero. During the analysis, we identified a number of "nonphysiological" pathways that carry flux *in silico* despite being unlikely to be activated *in vivo*; we provide detailed rationale and *in silico* implementation for these reactions in Supplementary text Section S1, Table S1.

2.4. Significant growth-coupled production

After implementing design conditions and knockouts in simulations, cell growth was maximized. In the M-model, to account for possible alternative molecule production, target growth-coupled production was minimized at maximum growth rate. In a ME-model, the flux distribution at the optimal growth rate is unique. When the substrate uptake rate is unconstrained, a ME-model is simulated under "batch growth mode" and the flux distribution at the maximal growth rate for central carbon metabolism was shown to be unique (O'Brien et al., 2013). Because of the additional constraints on gene expression, non-unique solutions can only exist if there exists an alternative pathway that had an equivalent protein as well as additional cell machinery synthesis costs. This coincidence is unlikely due to the difference in proteins construction cost (amino acid sequence differences) and turnover rates for each protein. Growth-coupled production level was evaluated in carbon yield determined by the following equation:

$$y_{product} = \frac{v_{product} * numC_{product}}{v_{substrate} * numC_{substrate}} \left(\frac{mmol\ carbon\ in\ product}{mmol\ carbon\ in\ substrate} \right) \quad (1)$$

where $y_{product}$ is target molecule carbon yield, $v_{product}$ and $v_{substrate}$ are target molecule growth-coupled production rate and substrate uptake rate in mmol gDW^{-1} h^{-1} respectively, $numC_{product}$ and $numC_{substrate}$ are number of carbons in target molecule and substrate molecule respectively.

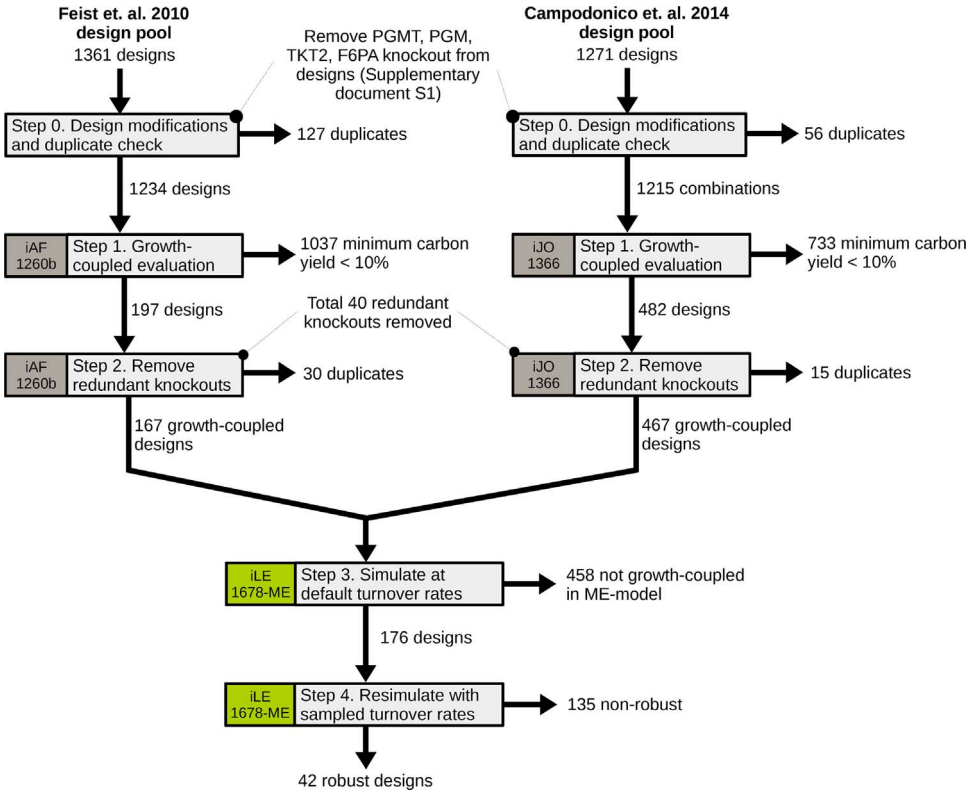

Fig. 2. Filtration workflow results. A pool of 2632 *unfiltered* designs were collected from previous studies and among them 634 had *significant growth-coupled production* and 42 had kinetically *robust* growth-coupled production. After Step 2, 42 redundant knockouts were removed. Designs were filtered for *significant growth-coupled production* (> 10% carbon yield) and subsequently robustness, *i.e.,* the ability to maintain growth-coupled production with various sets of sampled k_{eff}s.

A design was considered to have *significant growth-coupled production* if the flux through the biomass objective function was greater than 0.1 h^{-1} and $y_{product}$ was greater than 10%. The justification for using the cutoff of 10% carbon yield is as follows. First, there were many non-zero but extremely low yield (*e.g.* 0.1%) designs identified by the algorithm. It was imperative to filter these designs. Second, even though some compounds have higher theoretical yield than others, if a compound's theoretical yield is low, it was assumed to be uneconomical to produce given the target compounds. Third, both the optimal designs, as well as the suboptimal ones were kept to enable alternative design options for *in vivo* implementation (*i.e.,* as long as the 10% cutoff criterion was met for a design). Thus, a relatively low but "significant" level of 10% carbon yield was selected as a cutoff to satisfy the three factors above. A list of byproducts selected for analysis based on the previous work and their growth-coupled production status was given in Supplementary text, Section S4. The details of the designs with *significant growth-coupled production* are provided in Supplementary data.

2.5. Removal of redundant knockouts

A procedure was implemented to scan through each strain design and remove excessive reaction knockouts in order to minimize time for *in vitro* strain construction. For every design, knockouts were iteratively removed, and if a removal of a knockout from the design did not decrease any of three key performance criteria then the reaction was removed from the design. The three performance criteria we considered were adapted from Feist et al. (2010): carbon yield ($y_{product}$), maximum growth-rate without target molecule production ($\mu_{max,Remove\ Production}$ or $\mu_{max,RP}$), and substrate-specific productivity (*SSP*) (Supplementary Text, Section S2). As a cutoff for knockout filtration, removal of a redundant knockout did not drop $y_{product}$ by 1% carbon yield, $\mu_{max,RP}$ by 1% of original value, and *SSP* by 1% of original value.

2.6. Robustness evaluation of growth-coupled production with kinetic parameter sampling in the ME-model

Default enzyme turnover rates in the ME-model of *E. coli* have been determined by an omics-data driven optimization procedure that was described recently (Ebrahim et al., 2016). However, these turnover values can change between conditions, and there are many sources of uncertainty and error when determining kinetic parameters. Therefore, in this study a parameter robustness approach was implemented. Kinetic parameters were modified to decrease efficiency of the target molecule production pathway and increase efficiency of the competitive molecules production pathways which could render target molecule growth-coupled production pathway suboptimal. First, growth-coupled production was checked with the second optimal competitive compound production. If the design was *robust* against kinetic parameter change with respect to single competitive compound production, a more elaborate procedure was used to find all competitive compounds and then growth-coupled production was checked with all possible competitive compounds. Thus, growth-coupled production was considered to be *robust* if the kinetic change in the target and competitive pathways did not bring about an alternative production phenotype. The second optimal competitive molecule was identified in ME-model by re-simulation after removing the exchange reaction of the target molecule from the model. Third optimal competitive compound productions were found by removing target compound production and second optimal competitive compound production, and so on. Next, the sampled space was discretized into points at which pathway's turnover rates magnitude was adjusted and ME-model was re-simulated. Average magnitude of enzyme turnover rates of target and competitive molecules production pathway were sampled (*i.e.* average magnitude was increased or decreased) hence the sampled space was two-dimensional. If we found a simulation in which the cell could grow (represented by growth rate > 0.1 h^{-1}) without target molecule produc-

tion, the design was not *robust* (Supplementary text, Section S3). The details of the designs with *robust* growth-coupled production are provided in Supplementary data.

Possible enzyme turnover rates (k_{eff}s) were determined to range from 10^{-2} to 10^{6} s^{-1} based on a previous analysis of all enzyme turnover rates in the BRENDA database (Bar-even et al., 2011). Enzyme turnover rate magnitude was sampled between 10^{-2} s^{-1} and the default value for target pathways and between the default value and 10^{6} s^{-1} for the competitive pathways. Target molecule production was the first optimal production if the design was growth-coupled. Regarding the number of k_{eff} combinations being used, all designs passing through Step 2 of the workflow (634 designs) were tested with at least the default set of k_{eff}s. Depending on whether the competitive compound and target compound were precursors, three (precursor case) or nine (branched case) sets of k_{eff}s were used to test the growth-coupled production. More than nine sets of k_{eff}s were tested for some designs which need finer grain discretization (Supplementary text, Fig. S4). Specific cases were illustrated in the Supplementary text, Section S3, Fig. S3, and Fig. S4. Regarding the individual values of k_{eff}s, because all k_{eff}s of the enzymes in the pathway were scaled up or down by the same number, the relative fold difference between the enzymes in the same pathway is constant.

2.7. Data and reproducibility

Design specifications and relevant simulation results (*e.g.,* growth rate and compound production yield) were provided in the Supplementary data. Corresponding to the spreadsheets were four pools of design, including initial pool (before Step 0), M-model growth-coupled pool (after Step 2), ME-model growth-coupled pool (after Step 3), and ME-model robust pool (after Step 4). The *E. coli* M-models *i*AF1260b and *i*JO1366 can be downloaded from the BiGG database (King et al., 2016). The *E. coli* ME-model can be reconstructed using COBRAme package (Lloyd et al., 2018). M-model simulation was performed with COBRApy package (Ebrahim et al., 2013). ME-model simulation was performed with COBRApy, COBRAme, and solveME packages (Ebrahim et al., 2013; Lloyd et al., 2018; Yang et al., 2016).

3. Results

3.1. Strain design evaluation workflow

Evaluation and filtration of *in silico* strain designs is a rational step before undertaking expensive and time-consuming *in vivo* strain building and testing. Thus, a workflow was developed to filter previous strain designs and identify *robust* growth-coupled strains using both M- and ME-models (Fig. 1A). The workflow was used to evaluate and filter 2632 *E. coli* K-12 MG1655 knockout strain designs from previous model-driven strain design search studies by Feist et al. (2010) and Campodonico et al. (2014) that generated these designs using M-models (Fig. 1A) and the algorithms (OptKnock (Burgard et al., 2003), OptGene (Patil et al., 2005), RobustKnock (Tepper and Shlomi, 2009), and GDLS (Lun et al., 2009)).

Strain designs were first evaluated with the *E. coli* M-model that was used in the original study; thus strain designs identified in Feist et al. (2010) were evaluated in *i*AF1260b (Feist et al., 2007) and strain designs identified in Campodonico et al. (2014) were evaluated in *i*JO1366 (Orth et al., 2011). Here, a design was defined as a unique set of reaction knockouts, target molecule, substrate, oxygenation state, and a heterologous pathway (if present). Utilizing the workflow, 2632 *unfiltered* designs from the previous studies were first filtered to 634 *significant production* designs in the M-model (Fig. 2, steps 1 and 2), then further filtering led to a pool of 42 *robust* designs using the ME-model (Fig. 2, steps 3 and 4). The specific criteria used to filter the design pool is detailed below.

The reduction of the starting *unfiltered* strain design pool using the workflow in Fig. 2 can be rationalized by considering the magnitude of growth-coupled production, redundant knockouts, and robustness of production under kinetic parameters sampling. For Step 1, 2632 *unfiltered* designs were filtered for *significant production* (> 10% carbon yield). For Step 2, as the strain design algorithms OptKnock, RobustKnock, and GDLS used in Feist et al. (2010) and Campodonico et al. (2014) did not strictly include a penalty for knocking out extra reactions, a penalty was included here to look at a more restricted set. The impact of this filter can be understood in the example that a set of three useful knockouts is an equally optimal solution to that set plus an additional knockout of a reaction carrying no flux. A practical example from Campodonico et al. (2014), was that the single knockout of pyruvate-formate lyase (PFL) was sufficient for five designs with their target compounds being 1-propanol, 2-propanol, and 1,3-propanediol in the original design. Specifically for the design pool collected from previous work, the algorithms were run without a knockout penalty for OptKnock with three and five knockouts in Feist et al. (2010), RobustKnock with three knockouts in Campodonico et al. (2014), and GDLS with four knockouts in Campodonico et al. (2014). A notable exception is that OptGene was run with a knockout penalty and a maximum of ten knockouts in Feist et al. (2010). Additionally, after implementing a model modification and removing nonphysiological pathways (Supplementary text, Section S1), some knockouts might become obsolete. To address this issue, the second step, knockout filtration, was executed to remove knockouts that did not improve growth-coupled qualities significantly, and this step removed 40 knockouts from the collected design pool. Three factors that caused redundant knockouts to exist were the knockout and reaction directionality constraints from literature (Supplementary text, Section S1), the decision of the algorithm decision on adding the redundant knockouts to the design (no penalty for the redundant knockout), and marginal increase in production yield after applying the knockout (production yield improvement being non-zero but less than 1%). Determination of which of the three aforementioned factors caused the knockout to be redundant was performed for the designs that had their redundant knockouts removed. For Feist et al. (2010) study's designs, 68 of them were modified by Step 2. Among the 68 designs, new constraints implemented in the model made the knockout(s) redundant in 8 designs. Lack of a penalty for redundant knockouts added the redundant knockouts in 54 designs. Redundant knockouts which improved design production yields marginally were present in 6 designs. For Campodonico et al. (2014) study's designs, 94 of them were modified by Step 2. The number for the three factors were 86, 7, and 1, respectively. After steps 1 and 2 (Fig. 2), 634 *significant production* designs were obtained which had production of > 10% molar carbon yield and had their knockout redundancy checked. Redundant knockouts check was after *significant production* check because it was unnecessary to remove redundant knockouts for designs that were finally filtered out due to low production activity.

A final filtration step was performed on the growth-coupled design pool using a ME-model and varying kinetic parameters. In this part of the workflow, the ME-model was used as a tool to conservatively filter for the designs that met the growth-coupled criteria in both M and ME-models. As the ME-model's k_{eff}s are tunable, these parameters were selected to hinder growth-coupled production and the designs that were growth-coupled in the ME-model with multiple sets of k_{eff}s can be implemented *in vivo* with higher levels of confidence. Steps 3 and 4 in Fig. 2 (outlined in greater detail in Section 3.3 and Supplementary text Section S3) were used to filter designs using a ME-model. As ME-model simulations are more computationally expensive than M-model simulations, Steps 3 and 4 were arranged after M-model simulations. Ultimately, designs were meant to be filtered for their sustained growth-coupled production under both M and ME-model simulation, thus ME-model simulations were ordered last for economical computational resource usage. For step 3, designs' growth-coupled productions were checked using the default k_{eff}s set (Ebrahim et al., 2016). For

Table 2
Properties of strain designs. For each property, the left column referred to the pool of 634 *significant production* designs and the right column referred to the pool of 42 *robust* designs.

Target molecule	Number of designs		Max carbon yield (mmol C in product / mmol C in substrate)		Number of knockouts range	
	Significant production	Robust	Significant production	Robust	Significant production	Robust
1-Butanol	76	6	0.61	0.61	1–3	2–3
Isobutanol	17		0.63		1–3	
1-Propanol	58		0.55		1–4	
2-Propanol	10		0.41		2–4	
Ethanol	55		0.62		1–10	
1,4-Butanediol	191	17	0.61	0.54	1–4	2–4
2,3-Butanediol	2		0.44		3–3	
1,3-Propanediol	28		0.20		2–4	
Glycerol	1		0.34		3–3	
3-Hydroxyvalerate	0					
3-(R)-Hydroxybutyrate	9		0.60		2–4	
D-Lactate	63	6	0.96	0.95	2–10	5–10
Acrylate	33		0.50		1–4	
Acrylamide	14		0.50		2–4	
2-Oxopentanoate	0					
3-Methyl-2-oxobutanoate	0					
2-Oxobutanoate	0					
3-Hydroxypropanoate	29		0.80		2–4	
Fumarate	1		0.17		3–3	
Succinate	12	4	0.60	0.51	4–10	5–10
2-Oxoglutarate	1	1	0.12	0.14	5–5	5–5
Pyruvate	30	9	0.84	0.81	2–5	3–5
L-Alanine	4		0.95		3–5	
L-Glutamate	0					
L-Malate	0					
L-Serine	0					

step 4, designs which passed step 3 production criteria under the default k_{eff}s set were then checked with sampled k_{eff}s sets. Kinetic parameter sampling was unnecessary for designs that failed under the default k_{eff}s. After Step 4, 42 *robust* designs were obtained as the final *robust* set of growth-coupled designs.

3.2. Molecules with robust growth-coupled production in E. coli

Growth-coupled production of 19 molecules was identified from the set of 634 *significant production* designs from the M-model (Table 2). The majority of designs were alcohols (69% of *significant production* designs) that are fermented to recycle reduced cofactors generated by glycolysis. D-lactate and L-alanine were also terminal electron acceptors. The production of the electron accepting molecules including acrylate, acrylamide, 3-hydroxypropanoate, and 3-(R)-hydroxybutyrate (with an exception of anaerobic design co-producing ethanol) required oxygen as a co-acceptor. Some 1,4-butanediol designs also required oxygen as a co-acceptor. TCA cycle molecules (fumarate, succinate, and 2-oxoglutarate) were co-produced with alcohols. Pyruvate is not a terminal electron acceptor, and these designs required oxygen (Fig. 3). *Significant production* designs were found for all substrates (49% of designs were glucose, 33% xylose, and 18% glycerol) and oxygenation states (40% of designs were aerobic, 24% in ECOM, and 35% in anaerobic). Designs with glycerol as a substrate and anaerobic conditions favored alcohol designs (89% of glycerol designs and 80% of anaerobic designs produced alcohols). Most of the designs in the ECOM strain, due to its “aerobic fermentation” phenotype, produced molecules required oxygen as an electron co-acceptor (80% of ECOM designs successfully utilized oxygen as electron co-acceptor) (Fig. 4).

Growth-coupled production pathways were discussed and compared on a pathway level. The illustration of this discussion is shown in Fig. 3. For glucose or xylose as the substrate, the first divergence was the two glycolytic pathways Embden-Meyerhof-Parnas and Entner-Doudoroff and the pentose phosphate pathway (PPP). Design algorithms knocked out reaction(s) in glycolysis pathways or PPP in ~40% of the growth-coupled designs (Fig. 5A). The next divergence was at dihydroxyacetone phosphate (DHAP). DHAP is the precursor for three carbon alcohol production. Glycerol was converted into dihydroxyacetone to enter the second half of glycolysis. Reaction(s) consuming DHAP were targeted in ~20% of the designs. After glycolysis, the next divergence at phosphoenolpyruvate (PEP) branches to pyruvate or oxaloacetate and reaction(s) consuming PEP were targeted in ~15% of the designs (Fig. 5A). Pyruvate was an intermediate to a wide range of final products. If converted to acetyl-CoA, four possible final products excluding acrylate were ethanol, 1-butanol, 1,4-butanediol, and 3-(R)-hydroxybutyrate. Otherwise, the other four final products excluding acrylate were L-alanine, D-lactate, 2,3-butanediol, and isobutanol. Oxaloacetate, a compound in the TCA cycle, was another intermediate to a wide range of final products. If converted to succinyl-CoA, four possible final products were acrylamide, 3-hydroxypropanoate, acrylate, and 1-propanol. Otherwise, the other four final products were 1,3-propanediol, fumarate, 2-oxoglutarate, and succinate.

3.3. Robust designs after kinetic parameter sampling in the ME-model

Designs with *significant growth-coupled production* in the M-model were subsequently tested in a ME-model in this study. First, designs were simulated in the ME-model with default k_{eff}s which were determined as described in Ebrahim et al. (2016). For Step 3, even without computationally expensive kinetic parameter sampling, a single ME-model simulation with default k_{eff}s was able to filter out 458 of 634 designs as they did not result in a production phenotype. In Step 4, when performing kinetic parameters sampling taking into account the alternative production of the growth-optimal competitive compound(s), 134 out of 176 designs switched their phenotypes to alternative production, leaving 42 designs. Thus, 42 out of 634 designs were found to be *robust* (Fig. 2).

A kinetic parameter sampling method was sought to apply generally to all knockout designs utilized in the developed workflow. Using the turnover rate range reported in Bar-even et al. (2011), possible k_{eff}s were discretized between its default value and the extremity of the range. When simulating with the ME-model, so called ‘competitive molecules’ were identified as suboptimal replacements to target

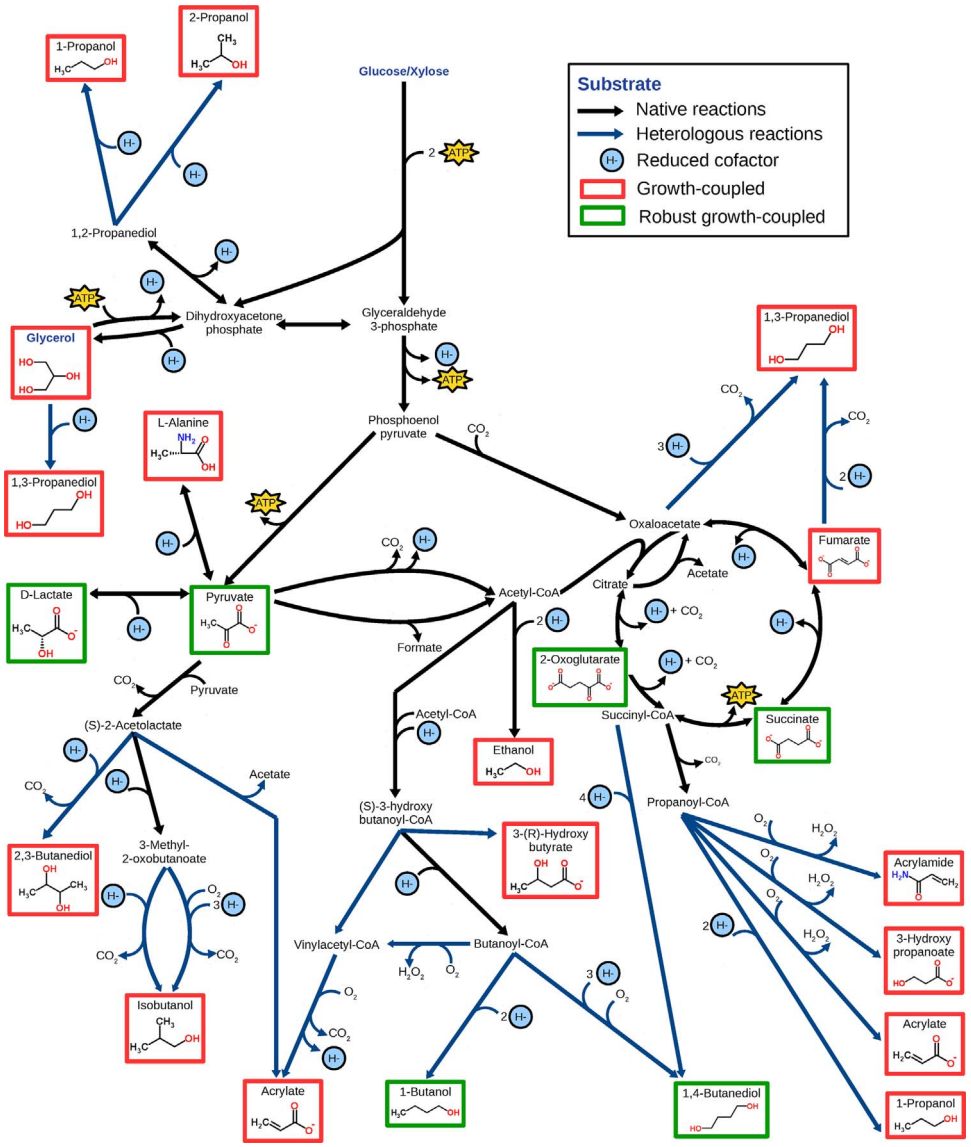

Fig. 3. Growth-coupled production pathway map. Target molecules, substrates, reduced cofactors (*e.g.*, NADH, NADPH), ATP, important branching points (*e.g.*, pyruvate, acetyl-CoA), and molecules contribute to carbon lost (*e.g.*, formate, CO_2) were annotated. Growth-coupled compounds in M-model only were red-boxed and robust compounds in both M and ME-model were green-boxed. Several pathways such as pentose phosphate pathway and Entner-Doudoroff were omitted to reduce figure complexity from network interconnectivity and emphasize production pathways.

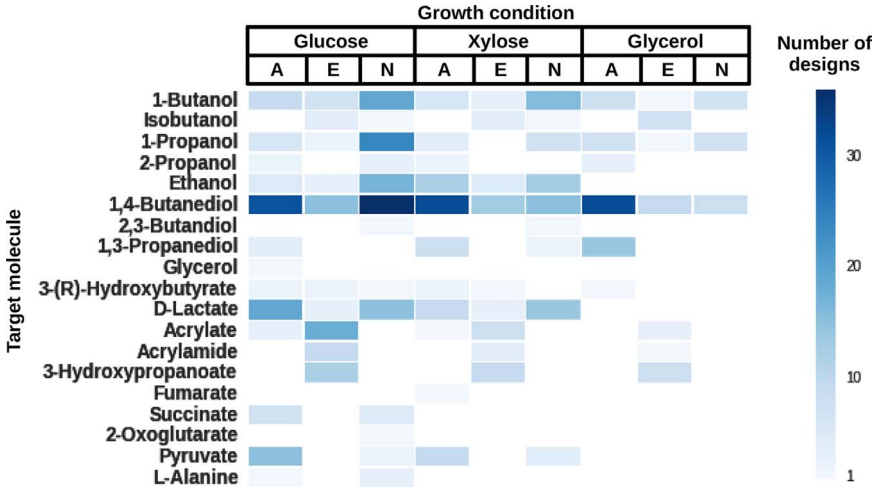

Fig. 4. Number of designs with *significant production* of target molecules in specific conditions. Oxygenation states included aerobic (A), ECOM (E), and anaerobic (N).

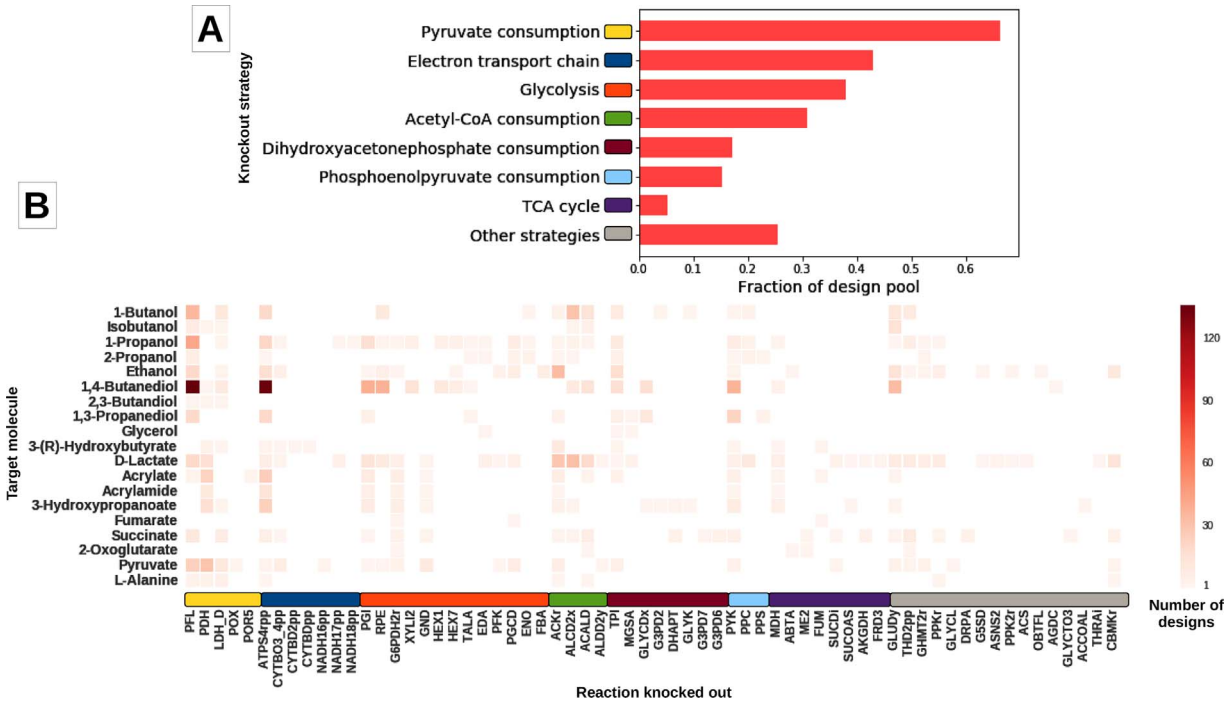

Fig. 5. Knockout strategy suggested by strain design search algorithm for *significant production designs*. (A) Fraction of growth-coupled design pool employed a specific knockout strategy. (B) Number of designs having a specific knockout for a target molecule. Knockout-to-strategy was mapped with colored patches between Figs. A and B.

molecules at default k_{eff} values, but shifted to being the optimally produced molecule under a different set of kinetic parameters altering the network enzyme efficiencies. Kinetic parameters were sampled in a way to break the target growth-coupled phenotype by making target molecule production pathways inefficient and/or competitive molecule production pathways more efficient. However, the growth-coupled production failure was more likely caused by the inefficient target molecule production (*i.e.,* higher competitive molecule production pathway's k_{eff}s; 68% of ME-model failed designs) rather than efficient competitive molecule production (*i.e.,* lower target molecule production pathway's k_{eff}s; 11% of ME-model failed designs). The majority of the failure modes being inefficient target molecule production pathways suggested that most of the strains' growth-coupled production had metabolic advantages, which were listed in Section 3.2, contributing to growth optimality besides protein costs. In other words, increasing enzyme efficiency or reducing the enzyme costs in the competitive molecule production pathway was only enough to raise the optimality of competitive molecule production sufficiently to replace growth-coupled production in 11% of the strain designs.

3.4. Knockout strategy and comparison of robust strain designs to in vitro production strains

Examining the *in silico significant growth-coupled production* design set, there were 64 knockouts suggested by the optimization-based strain design search algorithm that contribute to *significant growth-coupled production* in 634 strains. In this set, reactions around pyruvate are highly targeted (Fig. 5A), with pyruvate-formate lyase (PFL) being targeted the most. Electron transport chain reactions were the second most targeted group, and interestingly, ATP synthase was the second most knocked out and appeared in strain designs regardless of oxygenation conditions (ATPS4rpp, Fig. 5B). When comparing this set (*i.e., in silico* designs) to strain designs from a study which mined implemented designs from literature (*i.e., in vivo* designs) (King et al., 2017), similar reaction knockouts appeared frequently in both sets. Such shared reactions were in lactate dehydrogenase (LDH_D, *ldhA*), pyruvate-formate lyase (PFL, *pflB*), ethanol production (ALCD2x or ACALD, *adhE*), acetate production (ACKr or PTAr, *ackA* or *pta*), and succinate production (FRD3 or FUM or MDH or PPC, *frdA/B/C/D* or *fumABC* or *mdh or ppc*) (Fig. 5B). As an example for ME-model *robust* designs matching with *in vivo* production strains, 1,4-butanediol (14BDO) growth-coupled production strains were examined. Among 634 M-model growth-coupled strains, there were two groups of pathways for 14BDO production: one using a succinyl-CoA precursor and another using a butanoyl-CoA precursor (Fig. 6). However, after Steps 3 and 4, only the succinyl-CoA precursor group of pathways were *robust*. In those designs, knockouts of ethanol dehydrogenase, lactate dehydrogenase, and pyruvate-formate lyase were among the suggestion for *in silico* strains that matched with *in vivo* production strains with knockouts of *adhE*, *ldhA*, and *pflB* respectively (Yim et al., 2011).

In silico and *in vivo* implemented design sets were compared on a larger scale to determine the overlap in the reaction knockouts they contained. A design scenario was defined as a production objective for a selected media condition (substrate and oxygenation) and was used to select designs that matched the same scenario between the *in silico* and *in vivo* designs pools. Overall, 65 *in silico* (24/65 scenarios contained ME-model growth-coupled designs) and 32 *in vivo* design scenarios were compared and there were 12 common knockout combination scenarios overlapping (10/12 scenarios contained ME-model growth-coupled designs) when looking at the production objective and media conditions (Table S4). Non-overlapped *in silico* scenarios came from new target compounds (*e.g.,* acrylate) and varied media conditions (*e.g.,* xylose and glycerol, anaerobic and aerobic). Non-overlapped *in vivo* scenarios came from target compounds being out-of-scope for previous studies (*e.g.,* xylitol) or no growth-coupled designs were found for the scenarios (*e.g.,* no *in silico* designs were found for D-lactate in the glycerol and anaerobic condition). Between the *in vivo* and *in silico* design pools, there were no designs with identical knockout sets. However, there were common frequent knockouts which had the effect of blocking fermentative compound production. A major difference was that the *in vivo* designs commonly employed exhaustive fermentative compound production knockouts (*i.e.,* knockouts of all known fermentative). On the other hand, *in silico* designs used combination of several fermentative compound production knockouts and other non-intuitive

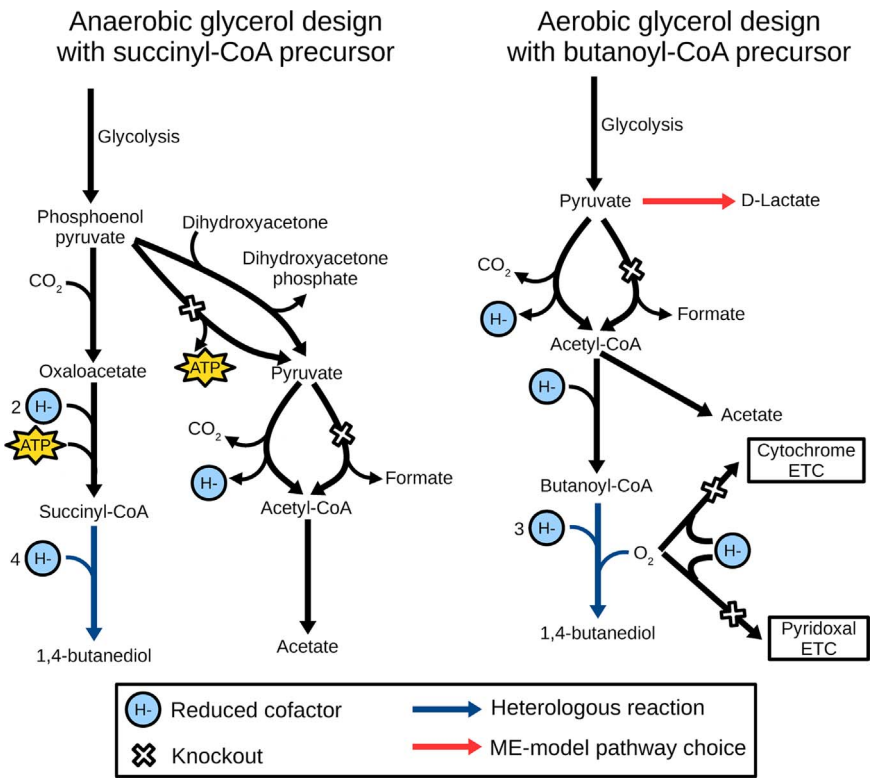

Fig. 6. Suggestion of succinyl-CoA precursor pathway by Metabolic-Expression model for *robust* 1,4-butanediol (14BDO) production pathway. ME-model simulation with kinetic parameter sampling revealed D-lactate production failure mode (red arrow) in butanoyl-CoA precursor pathway on the right. Maintaining production through all kinetic parameter sets, designs with succinyl-CoA precursor pathway were declared *robust*. ETC = electron transport chain.

knockouts such as knockouts in glycolysis (Table S4). Glycolysis knockout mutants such as Δ*pgi* and Δ*tpiA* were recently studied (Charusanti et al., 2010; Long et al., 2018; McCloskey et al., 2018) and glycolysis knockouts in strain designs were not widely considered in the *in vivo* implementation design set (Table S4). When considering how closely different strain design sets mapped to *in vivo* designs without the strict criteria of exact matching, designs with growth-coupled production confirmed by ME-model (after Step 3) were more closely resembled *in vivo* designs in 10 out of 12 scenarios than designs being filtered out (after Step 2 and filtered out by Step 3) (Fig. S5). In the remaining 2 scenarios of L-alanine (glucose, anaerobic) and ethanol (xylose, anaerobic) production, ME-model k_{eff}s restricted the *in silico* designs' growth-coupled production. Based on this comparison, ME-model verification was able to generate a reduced set of designs found to have more common knockouts to *in vivo* designs collected through the reference literature mining study.

4. Discussion

Genome-scale models that incorporate protein costs and kinetic parameters, such as the ME-model, are an advance in the scale and scope of microbial cell factory designs and enable more detailed studies. It was shown that next-generation models have better prediction capabilities (King et al., 2017) that could translate to increased confidence for prospective strain designs of growth-coupled production. Furthermore, the ME-model provides a mathematical framework to examine production states under a range of enzyme efficiencies (which are not known precisely *in vivo* (Davidi et al., 2016; Link et al., 2014)) by altering kinetic parameters. Thus, we deployed the use of the ME-model to filter pools of growth-coupled designs generated with previous approaches in order to find *robust* designs.

The workflow devised herein utilized both M- and ME-models to suggest high confidence strain designs with robust production, and it resulted in three pools of designs (with 634, 176, and 42 designs) with an increasing level of confidence, respectively, in their growth-coupled production based on the criteria laid out in this work. Knockout suggestions (Fig. 5B) were found to be in agreement with *in vitro* strains (King et al., 2017). A collection of 19 growth-coupled products consisting of alcohols, organic acids, and fatty acids in nine examined culturing conditions were found after evaluating strain designs from two previous large scale model-driven studies (Campodonico et al., 2014; Feist et al., 2010). The result from the two previous studies and this study showcase *E. coli*'s potential as a microbial cell factory for various relevant industrial products and reveals promising strain designs.

With the inclusion of kinetic parameters in the ME-model, there are three factors to be considered when performing the kinetic parameter sampling: performance metric, enzyme selection, and sampling range. In this work, the performance metric was growth-coupled production of the target molecule exceeding 10% carbon yield. The average magnitudes of k_{eff}s in target and competitive pathways were selected for sampling. To achieve the highest level of production confidence, the k_{eff}s were sampled from their default values to the upper bound of 10^6 s^{-1} or lower bound of 10^{-2} s^{-1} of BRENDA enzyme collection (Bar-even et al., 2011). Overall, the objective for ME-model simulations was growth-coupled production robustness and the sampling procedure was designed with strict criteria to raise the level of confidence for *robust* strain designs that passed the filter. With kinetic parameter sampling, the ME-model revealed how enzyme efficiency and pathway tradeoffs can affect growth-coupled production phenotypes.

Growth-coupled production is a desirable trait for microbial cell factories as production of the target molecule is necessary for cell growth. Because cell growth *in silico* was driven to optimal growth-coupled production among various possible phenotypes by maximizing flux through the biomass objective function, an analogous *in vivo* drive will be necessary to make *in vivo* implementation succeed. Adaptive Laboratory Evolution (ALE) is an *in vivo* platform to drive cells toward optimal growth and can thus be used to optimize production rate in

growth-coupled strains. Several studies have successfully used ALE to increase growth-coupled production (Baek et al., 2017; Fong et al., 2005; Zhang et al., 2007). ALE was also found to increase product (Atsumi et al., 2010; Mundhada et al., 2017) and culture supplement tolerance (Mohamed et al., 2017), favorable traits for production under high concentrations of target molecules. Besides implementing knock-outs, there exists other biological barriers that are beyond the scope of M- and ME-models, such as regulation, transporter engineering, and compound toxicity, that have been shown to affect molecule over-production (Davy et al., 2017). Regulation and transporter availability can influence growth rate and cell factory performance (Chubukov et al., 2016; Hong and Nielsen, 2012). As a continuation of the workflow (Fig. 1A), available modeling methods for regulation integration (Blazier and Papin, 2012) and transporter expression (Dunlop et al., 2011) could be integrated, and further design improvements could be suggested to increase the confidence in strain designs.

Acknowledgements

We would like to thank Colton J. Lloyd and Laurence Yang for their guidance and suggestion on the setup and performing simulation during the course of this project. This work was funded by the Novo Nordisk Foundation Grant Number NNF10CC1016517 and by grant CN-15-1420 from the University of California Institute for Mexico and the United States (UC MEXUS) and the Consejo Nacional de Ciencia y Techologica de Mexico (CONACYT).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.mec.2018.e00080.

References

Atsumi, S., Wu, T.-Y., Machado, I.M.P., Huang, W.-C., Chen, P.-Y., Pellegrini, M., Liao, J.C., 2010. Evolution, genomic analysis, and reconstruction of isobutanol tolerance in *Escherichia coli*. Mol. Syst. Biol. 6, 449. http://dx.doi.org/10.1038/msb.2010.98.

Baek, S.-H., Kwon, E.Y., Bae, S.-J., Cho, B.-R., Kim, S.-Y., Hahn, J.-S., 2017. Improvement of d-Lactic acid production in *Saccharomyces cerevisiae* under acidic conditions by evolutionary and rational metabolic engineering. Biotechnol. J. 12, 1700015. http://dx.doi.org/10.1002/biot.201700015.

Bar-even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D.S., Milo, R., Taw, D.S., Milo, R., 2011. The moderately efficient enzyme: evolutionary and physicochemical. Biochemistry 50, 4402–4410. http://dx.doi.org/10.1021/bi2002289.

Blazier, A.S., Papin, J.A., 2012. Integration of expression data in genome-scale metabolic network reconstructions. Front. Physiol. 3, 299. http://dx.doi.org/10.3389/fphys.2012.00299.

Burgard, A., Pharkya, P., Maranas, C., 2003. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. 84, 647–657. http://dx.doi.org/10.1002/bit.10803.

Campodonico, M.A., Andrews, B.A., Asenjo, J.A., Palsson, B.O., Feist, A.M., 2014. Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path. Metab. Eng. 25, 140–158. http://dx.doi.org/10.1016/j.ymben.2014.07.009.

Charusanti, P., Conrad, T.M., Knight, E.M., Venkataraman, K., Fong, N.L., Xie, B., Gao, Y., Palsson, B.Ø., 2010. Genetic basis of growth adaptation of Escherichia coli after deletion of pgi, a major metabolic gene. PLoS Genet. 6, e1001186. http://dx.doi.org/10.1371/journal.pgen.1001186.

Chubukov, V., Mukhopadhyay, A., Petzold, C.J., Keasling, J.D., Martín, H.G., 2016. Synthetic and systems biology for microbial production of commodity chemicals. npj Syst. Biol. Appl. 2, 16009. http://dx.doi.org/10.1038/npjsba.2016.9.

Clark, D.P., 1989. The fermentation pathways of *Escherichia coli*. FEMS Microbiol. Lett. 63, 223–234. http://dx.doi.org/10.1111/j.1574-6968.1989.tb03398.x.

Davidi, D., Noor, E., Liebermeister, W., Bar-Even, A., Flamholz, A., Tummler, K., Barenholz, U., Goldenfeld, M., Shlomi, T., Milo, R., 2016. Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements. Proc. Natl. Acad. Sci. 113, 3401–3406. http://dx.doi.org/10.1073/pnas.1514240113.

Davy, A.M., Kildegaard, H.F., Andersen, M.R., 2017. Cell factory engineering. Cell Syst. 4, 262–275. http://dx.doi.org/10.1016/j.cels.2017.02.010.

Dunlop, M.J., Dossani, Z.Y., Szmidt, H.L., Chu, H.C., Lee, T.S., Keasling, J.D., Hadi, M.Z., Mukhopadhyay, A., 2011. Engineering microbial biofuel tolerance and export using efflux pumps. Mol. Syst. Biol. 7, 487. http://dx.doi.org/10.1038/msb.2011.21.

Ebrahim, A., Brunk, E., Tan, J., O'Brien, E.J., Kim, D., Szubin, R., Lerman, J.A., Lechner, A., Sastry, A., Bordbar, A., Feist, A.M., Palsson, B.O., 2016. Multi-omic data integration enables discovery of hidden biological regularities. Nat. Commun. 7, 13091. http://dx.doi.org/10.1038/ncomms13091.

Ebrahim, A., Lerman, J.A., Palsson, B.O., Hyduke, D.R., 2013. COBRApy: constraints-based reconstruction and analysis for python. BMC Syst. Biol. 7, 74. http://dx.doi.org/10.1186/1752-0509-7-74.

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3, 121. http://dx.doi.org/10.1038/msb4100155.

Feist, A.M., Zielinski, D.C., Orth, J.D., Schellenberger, J., Markus, J., Palsson, B.Ø., 2010. Model-driven evalution of the production potential for growth coupled products of Escherichia coli. Metab. Eng. 12, 173–186. http://dx.doi.org/10.1016/j.ymben.2009.10.003.Model-driven.

Fong, S.S., Burgard, A.P., Herring, C.D., Knight, E.M., Blattner, F.R., Maranas, C.M., Palsson, B.O., 2005. In silico design and adaptive evolution of Escherichia coli for production of lactic acid. Biotechnol. Bioeng. 91, 643–648.

Hong, K.K., Nielsen, J., 2012. Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. Cell. Mol. Life Sci. 69, 2671–2690. http://dx.doi.org/10.1007/s00018-012-0945-1.

Johnson, F.X., 2008. Industrial biotechnology and biomass utilisation. Futur. Prospect. Ind. Biotechnol.

King, Z.A., Lu, J., Ager, A.D., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A., Palsson, B.O., Lewis, N.E., 2016. BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res. 44. http://dx.doi.org/10.1093/nar/gkv1049.

King, Z.A., O'Brien, E.J., Feist, A.M., Palsson, B.O., 2017. Literature mining supports a next-generation modeling approach to predict cellular byproduct secretion. Metab. Eng. 39, 220–227. http://dx.doi.org/10.1016/j.ymben.2016.12.004.

Lee, S.Y., Kim, H.U., 2015. Systems strategies for developing industrial microbial strains. Nat. Biotechnol. 33, 1061–1072. http://dx.doi.org/10.1038/nbt.3365.

Lewis, N.E., Nagarajan, H., Palsson, B.O., 2012. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat. Rev. Microbiol. 10, 291–305. http://dx.doi.org/10.1038/nrmicro2737.

Link, H., Christodoulou, D., Sauer, U., 2014. Advancing metabolic models with kinetic information. Curr. Opin. Biotechnol. 29, 8–14. http://dx.doi.org/10.1016/j.copbio.2014.01.015.

Liu, J.K., O'Brien, E.J., Lerman, J.A., Zengler, K., Palsson, B.O., Feist, A.M., 2014. Reconstruction and modeling protein translocation and compartmentalization in *Escherichia coli* at the genome-scale. BMC Syst. Biol. 8, 110. http://dx.doi.org/10.1186/s12918-014-0110-6.

Lloyd, C.J., Ebrahim, A., Yang, L., King, Z.A., Catoiu, E., O'Brien, E.J., Liu, J.K., Palsson, B.O., 2018. COBRAme: a computational framework for genome-scale models of metabolism and gene expression. PLOS Comput. Biol. 14, e1006302. http://dx.doi.org/10.1371/journal.pcbi.1006302.

Long, C.P., Gonzalez, J.E., Feist, A.M., Palsson, B.O., Antoniewicz, M.R., 2018. Dissecting the genetic and metabolic mechanisms of adaptation to the knockout of a major metabolic enzyme in *Escherichia coli*. Proc. Natl. Acad. Sci. 115, 222–227. http://dx.doi.org/10.1073/pnas.1716056115.

Lun, D.S., Rockwell, G., Guido, N.J., Baym, M., Kelner, J.A., Berger, B., Galagan, J.E., Church, G.M., 2009. Large-scale identification of genetic design strategies using local search. Mol. Syst. Biol. 5, 296. http://dx.doi.org/10.1038/msb.2009.57.

Ma, D., Saunders, M.A., 2014. Solving multiscale linear programs using the simplex method in quadruple precision, In: Numerical Analysis and Optimization III, Springer Proceedings in Mathematics & Statistics. Springer, Cham, pp. 223–235.

Machado, D., Herrgård, M.J., 2015. Co-evolution of strain design methods based on flux balance and elementary mode analysis. Metab. Eng. Commun. 2, 85–92. http://dx.doi.org/10.1016/j.meteno.2015.04.001.

Maia, P., Rocha, M., Rocha, I., 2016. *In silico* constraint-based strain optimization methods: the quest for optimal cell factories. Microbiol. Mol. Biol. Rev. 80, 45–67. http://dx.doi.org/10.1128/MMBR.00014-15.

McCloskey, D., Xu, S., Sandberg, T.E., Brunk, E., Hefner, Y., Szubin, R., Feist, A.M., Palsson, B.O., 2018. Adaptation to the coupling of glycolysis to toxic methylglyoxal production in tpiA deletion strains of Escherichia coli requires synchronized and counterintuitive genetic changes. Metab. Eng. 48, 82–93. http://dx.doi.org/10.1016/j.ymben.2018.05.012.

Mohamed, E.T., Wang, S., Lennen, R.M., Herrgård, M.J., Simmons, B.A., Singer, S.W., Feist, A.M., 2017. Generation of a platform strain for ionic liquid tolerance using adaptive laboratory evolution. Microb. Cell Fact. 16, 204. http://dx.doi.org/10.1186/s12934-017-0819-1.

Mundhada, H., Seoane, J.M., Schneider, K., Koza, A., Christensen, H.B., Klein, T., Phaneuf, P.V., Herrgard, M., Feist, A.M., Nielsen, A.T., 2017. Increased production of L-serine in Escherichia coli through Adaptive Laboratory Evolution. Metab. Eng. 39, 141–150. http://dx.doi.org/10.1016/j.ymben.2016.11.008.

O'Brien, E.J., Lerman, J.A., Chang, R.L., Hyduke, D.R., Palsson, B., 2013. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol. Syst. Biol. 9, 693. http://dx.doi.org/10.1038/msb.2013.52.

Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B., 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011. Mol. Syst. Biol. 7, 535. http://dx.doi.org/10.1038/msb.2011.65.

Park, J.H., Lee, S.Y., 2008. Towards systems metabolic engineering of microorganisms for amino acid production. Curr. Opin. Biotechnol. 19, 454–460. http://dx.doi.org/10.1016/j.copbio.2008.08.007.

Patil, K.R., Rocha, I., Forster, J., Nielsen, J., 2005. Evolutionary programming as a platform for *in silico* metabolic engineering. BMC Bioinforma. 6, 308. http://dx.doi.org/10.1186/1471-2105-6-308.

Portnoy, V.A., Herrgård, M.J., Palsson, B.Ø., 2008. Aerobic fermentation of D-glucose by an evolved cytochrome oxidase-deficient *Escherichia coli* strain. Appl. Environ. Microbiol. 74, 7561–7569. http://dx.doi.org/10.1128/AEM.00880-08.

Sittig, M., Weil, B.H., 1954. Raw materials for chemicals from petroleum, In: Literature Resources, Advances in Chemistry. American Chemical Society, pp. 36–327. ⟨http://dx.doi.org/10.1021/ba-1954-0010.ch036⟩.

Tepper, N., Shlomi, T., 2009. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics 26, 536–543. http://dx.doi.org/10.1093/bioinformatics/btp704.

Thiele, I., Palsson, B., 2010. A protocol for generating a high quality genome-scale metabolic reconstruction. Nat. Protoc. 5, 93–121.

Yang, L., Ma, D., Ebrahim, A., Lloyd, C.J., Saunders, M.A., Palsson, B.O., 2016. solveME: fast and reliable solution of nonlinear ME models. BMC Bioinform. 17, 1–9. http://dx.doi.org/10.1186/s12859-016-1240-1.

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J.D., Osterhout, R.E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H.B., Andrae, S., Yang, T.H., Lee, S.Y., Burk, M.J., Van Dien, S., 2011. Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. Nat. Chem. Biol. 7, 445–452. http://dx.doi.org/10.1038/nchembio.580.

Zhang, X., Jantama, K., Moore, J.C., Shanmugam, K.T., Ingram, L.O., 2007. Production of L-alanine by metabolically engineered *Escherichia coli*. Appl. Microbiol. Biotechnol. 77, 355–366. http://dx.doi.org/10.1007/s00253-007-1170-y.